先端人工知能論
生成モデル

2017.10.17

佐藤一誠

もくじ

授業で使う記法

x_i ：i番目の観測データ (i=1,…,n)

$x_{i,d}$ ：x_iのd次元要素 (d=1,…,D)

$x_{1:n}$ ：n個の観測データ

z_i ：i番目の潜在変数 (i=1,…,n)

$z_{i,k}$ ：z_iのd次元要素 (k=1,…,K)

$z_{1:n}$ ：n個の潜在変数

$p(x|\theta)$: θをパラメータとする確率分布

$p_\theta(x)$と書くこともある

x = 1 となる確率をπ，x = 0 となる確率を1-πとする．
二値確率変数xが，確率分布

$$\mathrm{Be}(x \mid \pi) = \pi^{x}(1-\pi)^{1-x}$$

に従うとき，x はπをパラメータとする
ベルヌーイ分布に従うという

実数値確率変数xが，確率密度関数

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right)$$

をもつとき，xは平均μ，分散σ^2をパラメータ
とする正規分布に従うという

識別モデルと生成モデルの学習

識別モデル

学習データ

$$\{(y_i, x_i)\}_{i=1}^n \Rightarrow y_i = f_\theta(x_i)$$

関数を学習

生成モデル

学習データ:

$$\{x_i\}_{i=1}^n \Rightarrow x_i = f_\theta(z_i)$$

関数及び潜在変数を学習

識別モデル[illegible]習

生成モデルは統計モデルによって
記述されるため、実際には

$$x_i \sim \underline{p_\theta}(x \mid \underline{z_i})$$

確率分布及び潜在変数を学習

生[illegible]

学習データ:

$$\{x_i\}_{i=1}^n$$

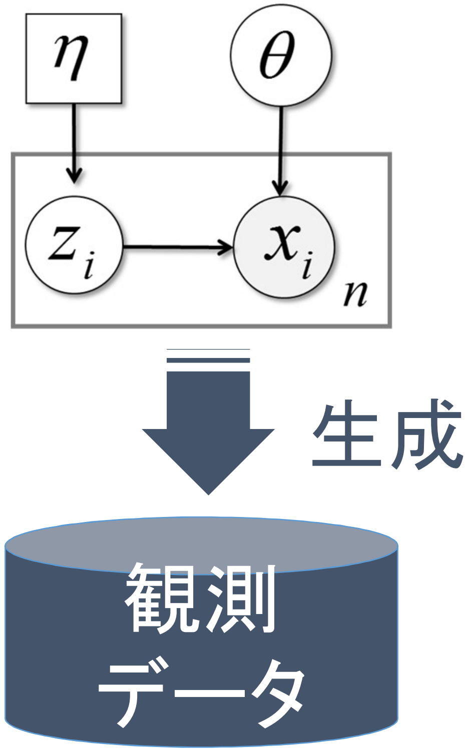
⇒

$$x_i = \underline{f_\theta}(\underline{z_i})$$

関数及び潜在変数を学習

生成モデル

データの生成過程を数理モデルにより表現したもの



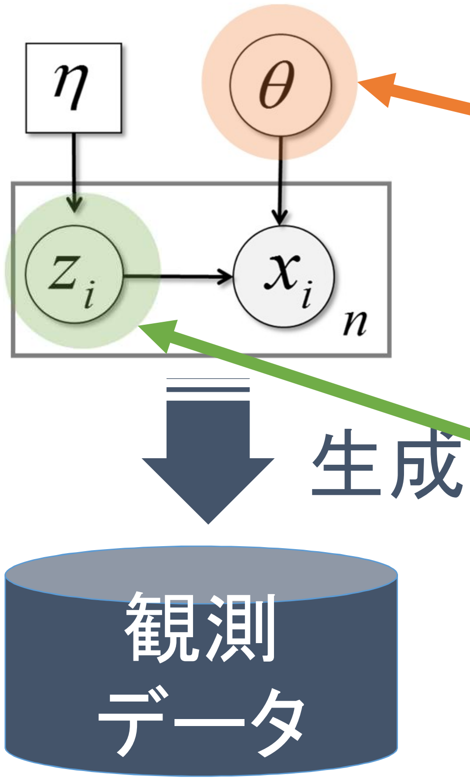
データを生成した
数理モデル
（統計モデル）

学習

観測データは
ある数理モデル
から生成されたと仮定

生成モデル

データの生成過程を数理モデルにより表現したもの



パラメータ:
データに共通する部分

潜在変数:
データに固有の部分

変分推論
(Variational inference)

目的関数

$$\log p_\theta(x_i)$$

$$= \log \int p_\theta(x_i, z_i) dz_i$$

何ら仮定をおかずに
確率分布を導入

$$= \log \int q(z_i) \frac{p_\theta(x_i, z_i)}{q(z_i)} dz_i$$

目的関数

$$\log p_\theta(x_i)$$

$$= \log \int p_\theta(x_i, z_i) dz_i$$

何ら仮定をおかずに
確率分布を導入

$$= \log \int q(z_i) \frac{p_\theta(x_i, z_i)}{q(z_i)} dz_i$$

$$= \log E_{q(z_i)}\left[\frac{p_\theta(x_i, z_i)}{q(z_i)} \right]$$

目的関数

$$\log p_\theta(x_i)$$

$$= \log \int p_\theta(x_i, z_i) dz_i$$

何ら仮定をおかずに
確率分布を導入

$$= \log \int q(z_i) \frac{p_\theta(x_i, z_i)}{q(z_i)} dz_i$$

$$= \log E_{q(z_i)}\left[\frac{p_\theta(x_i, z_i)}{q(z_i)}\right]$$

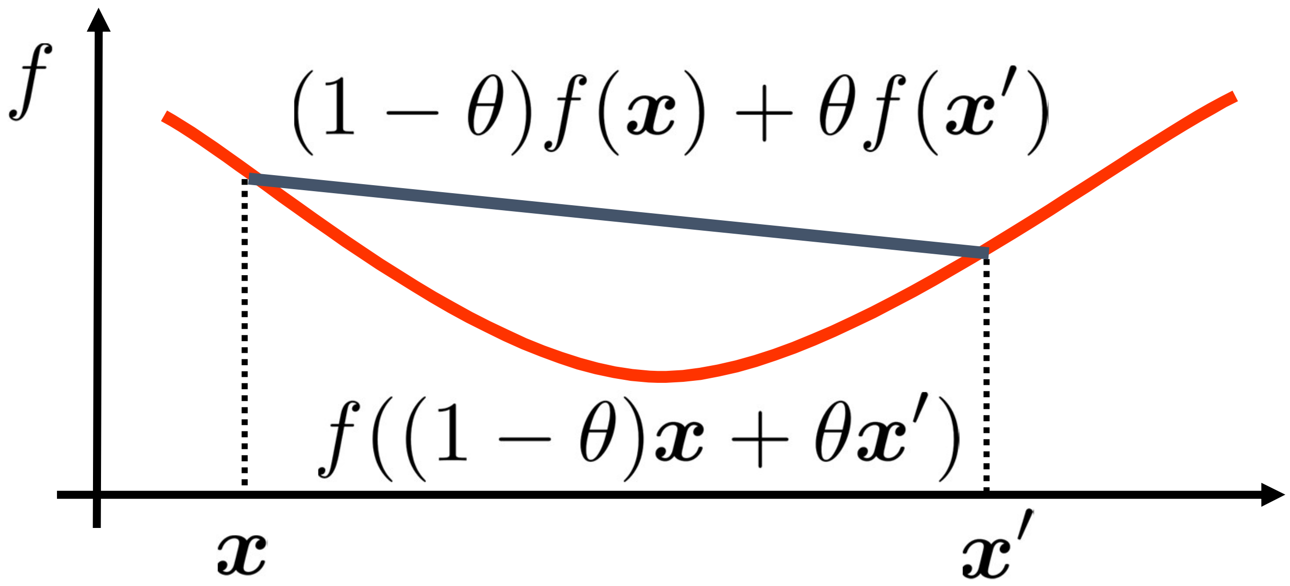
$$\geq E_{q(z_i)}\left[\log \frac{p_\theta(x_i, z_i)}{q(z_i)}\right]$$

凸関数

- 任意の $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ 任意の$\theta \in [0, 1]$に対して,

$$f((1-\theta)x+\theta x') \leq (1-\theta) f(x) + \theta f(x')$$

ならば, f を凸関数(convex function)とよぶ

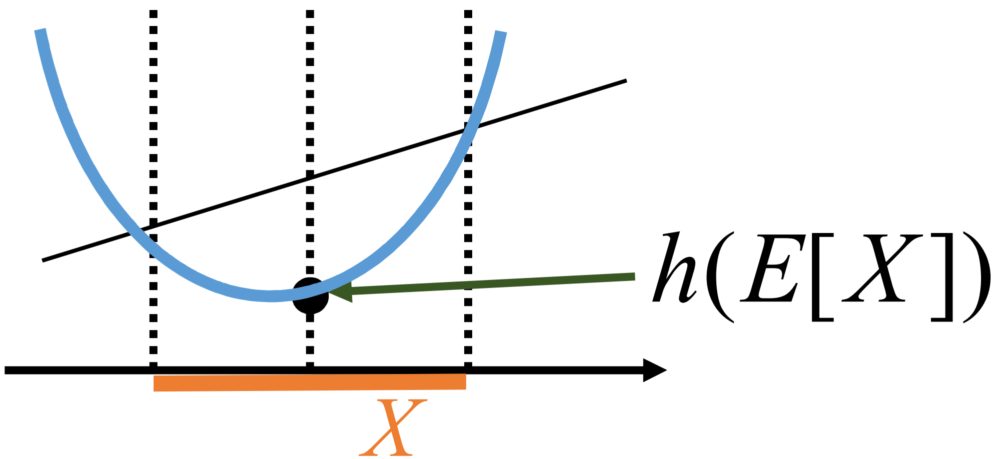


Jensenの不等式

$h(\boldsymbol{x})$ を凸関数とする.
確率変数Xに対して

$$h(E[X]) \leq E[h(X)]$$

が成り立つ.



目的関数

$$\log p_\theta(x_i)$$

$$= \log \int p_\theta(x_i, z_i) dz_i$$

何ら仮定をおかずに
確率分布を導入

$$= \log \int q(z_i) \frac{p_\theta(x_i, z_i)}{q(z_i)} dz_i$$

$$= \log E_{q(z_i)}\left[\frac{p_\theta(x_i, z_i)}{q(z_i)}\right]$$

$$\geq E_{q(z_i)}\left[\log \frac{p_\theta(x_i, z_i)}{q(z_i)}\right]$$

目的関数

$$\log p_\theta(x_i)$$

$$= \log \int p_\theta(x_i, z_i) dz_i$$

何ら仮定をおかずに
確率分布を導入

$$= \log \int q(z_i) \frac{p_\theta(x_i, z_i)}{q(z_i)} dz_i$$

$$\geq \int q(z_i) \log \frac{p_\theta(x_i, z_i)}{q(z_i)} dz_i$$

目的関数

$$\log p_\theta(x_i)$$

$$\geq \int q(z_i) \log \frac{p_\theta(x_i, z_i)}{q(z_i)} dz_i$$

$$\equiv L[q(z_i), \theta; x_i]$$ 変分下限などと呼ばれる

変分下限最大化へ帰着

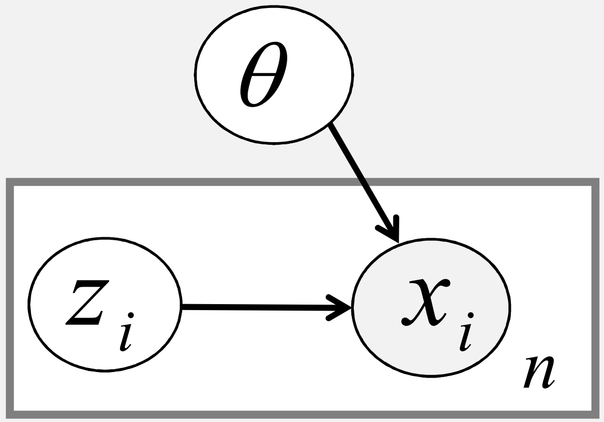
$$(q^*(z_i), \theta^*) = \underset{q(z_i),\theta}{\arg\max}\, L[q(z_i), \theta; x_i]$$

練習問題　データが複数ある場合

$\log p_{\theta}(x_{1:n})$ の変分下限が

$$\log p_{\theta}(x_{1:n}) \geq \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$

となることを示せ



解答

$$\log p_\theta(x_{1:n})$$

$$= \log \int p_\theta(x_{1:n}, z_{1:n}) dz_{1:n}$$

何ら仮定をおかずに
確率分布を導入

$$= \log \int q(z_{1:n}) \frac{p_\theta(x_{1:n}, z_{1:n})}{q(z_{1:n})} dz_{1:n}$$

$$\geq \int q(z_{1:n}) \log \frac{p_\theta(x_{1:n}, z_{1:n})}{q(z_{1:n})} dz_{1:n}$$

$$= \int q(z_{1:n}) (\log p_\theta(x_{1:n}, z_{1:n}) - \log q(z_{1:n})) dz_{1:n}$$

解答

$$\int q(z_{1:n}) \log p_\theta(x_{1:n}, z_{1:n}) dz_{1:n}$$

$$= \int \prod_{i=1}^{n} q(z_i) \log \prod_{i=1}^{n} p_\theta(x_i, z_i) dz_1 dz_2 \cdots dz_n$$

$$= \int \prod_{i=1}^{n} q(z_i) \sum_{i=1}^{n} \log p_\theta(x_i, z_i) dz_1 dz_2 \cdots dz_n$$

解答

$$\int q(z_1)q(z_2)(\log p_\theta(x_1, z_1) + \log p_\theta(x_2, z_2))dz_1 dz_2$$

$$= \int q(z_1)q(z_2)\log p_\theta(x_1, z_1)dz_1 dz_2$$

$$+ \int q(z_1)q(z_2)\log p_\theta(x_2, z_2)dz_1 dz_2$$

$$= \int q(z_1)\log p_\theta(x_1, z_1)dz_1 \int q(z_2)dz_2$$

$$+ \int q(z_2)\log p_\theta(x_2, z_2)dz_2 \int q(z_1)dz_1$$

目的関数

$$\underline{\log p_\theta(x_{1:n})}$$

$$\geq \sum_{i=1}^{n} \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$ <u>変分下限</u>などと呼ばれる

変分下限最大化へ帰着

$$(q^*(z_{1:n}), \theta^*) = \underset{q(z_{1:n}),\theta}{\arg\max} \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$

目的関数

$$\log p_\theta(x_{1:n})$$

$$\geq \sum_{i=1}^{n} \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$ 変分下限などと呼ばれる

交互最適化アルゴリズム

$$q_t(z_i) = \underset{q(z_i)}{\arg\max}\, L[q(z_i), \theta_{t-1}; x_i]$$

$$\theta_t = \underset{\theta}{\arg\max} \sum_{i=1}^{n} L[q_t(z_i), \theta; x_i]$$

目的関数

$$\underline{\log p_\theta(x_{1:n})}$$

$$\geq \sum_{i=1}^{n} \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$ 変分下限などと呼ばれる

交互最適化アルゴリズム

$$q_t(z_i) = \arg\max_{q(z_i)} L[q(z_i), \theta_{t-1}; x_i]$$

$$\theta_t = \arg\max_{\theta} \sum_{i=1}^{n} L[q_t(z_i), \theta; x_i]$$

目的関数

$$\log p_\theta(x_{1:n})$$

$$\geq \sum_{i=1}^{n} \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$ 変分下限などと呼ばれる

確率分布の最適化

$$q_t(z_i) = \underset{q(z_i)}{\arg\max}\, L[q(z_i), \theta_{t-1}; x_i]$$

⇒ 変分法で解ける

変分法

汎関数

関数を入力とする関数

例

$$L[f(x)] = \int_0^{\pi} f(x)dx$$

f(x)=sin(x), cos(x)を入力すると
Lの値が変わる

変分法

微分法

目的：関数$f(x)$ の極値を求めたい

$$\frac{df(x)}{dx} = 0$$ となるx を求める

変分法

目的：汎関数$L[f(x)]$ の極値を求めたい

$$\frac{\delta L[f]}{\delta f} = 0$$ となる関数$f(x)$ を求める

変分法

汎関数を以下のように定義する

$$L[q(x)] = \int f(x,q)dx$$

例 エントロピー

$$L[q(x)] = \int -q(x)\log q(x)dx$$

汎関数$L[q(x)]$ の極値を与える$q(x)$は以下のオイラー･ラグランジュ方程式を解けばよい

$$\frac{\partial f(x,q)}{\partial q(x)} = 0$$

目的関数

$$\log p_\theta(x_i)$$

$$\geq \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv L[q(z_i), \theta; x_i]$$

この部分を$q(z_i)$で
微分＝0を解けば良い

確率分布の最適化

$$q_t(z_i) = \arg\max_{q(z_i)} L[q(z_i), \theta_t; x_i]$$

➡ 変分法で解ける（演習）

目的関数

$$\log p_\theta(x_i)$$

$$\geq \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv L[q(z_i), \theta; x_i]$$

この部分を$q(z_i)$で微分＝0を解けば良い

確率分布の最適化

$$q_t(z_i) = p_{\theta_t}(z_i \mid x_i)$$

事後分布

導出

λを定数として，以下の汎関数の極値を求める

$$\int q(z_i)\log\frac{p_\theta(x_i\,|\,z_i)p(z_i)}{q(z_i)}dz_i+\underline{\lambda\left(1-\int q(z_i)dz_i\right)}$$

制約付き最適化

$$\partial_{q(z_i)}\left(q(z_i)\log\frac{p_\theta(x_i\,|\,z_i)p(z_i)}{q(z_i)}-\lambda q(z_i)\right)=0$$

$$\Leftrightarrow \log\frac{p_\theta(x_i\,|\,z_i)p(z_i)}{q(z_i)}-1-\lambda=0$$

導出

λを定数として，以下の汎関数の極値を求める

$$\int q(z_i)\log\frac{p_\theta(x_i\mid z_i)p(z_i)}{q(z_i)}dz_i+\underbrace{\lambda\left(1-\int q(z_i)dz_i\right)}_{\text{制約付き最適化}}$$

$$q_t(z_i)=e^{-\lambda-1}\,p_{\theta_t}(x_i\mid z_i)p(z_i)$$

$\int q_t(z_i)dz_i=1$ より

$$q_t(z_i)=\frac{p_{\theta_t}(x_i\mid z_i)p(z_i)}{\int p_{\theta_t}(x_i\mid z_i)p(z_i)dz_i}=p_{\theta_t}(z_i\mid x_i)$$

目的関数

$$\log p_\theta(x_{1:n})$$

$$\geq \sum_{i=1}^{n} \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$ 変分下限などと呼ばれる

交互最適化アルゴリズム

$$q_t(z_i) = \underset{q(z_i)}{\arg\max}\, L[q(z_i), \theta_t; x_i]$$

$$\theta_t = \underset{\theta}{\arg\max} \sum_{i=1}^{n} L[q_t(z_i), \theta; x_i]$$

目的関数

$$\log p_\theta(x_{1:n})$$

$$\geq \sum_{i=1}^{n} \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$ 変分下限などと呼ばれる

交互最適化アルゴリズム

$$q_t(z_i) = p_\theta(z_i \mid x_i)$$

EMアルゴリズムと呼ばれる

$$\theta_t = \arg\max_\theta \sum_{i=1}^{n} L[q_t(z_i), \theta; x_i]$$

目的関数

$$\log p_\theta(x_{1:n})$$

$$\geq \sum_{i=1}^{n} \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$ 変分下限などと呼ばれる

交互最適化アルゴリズム

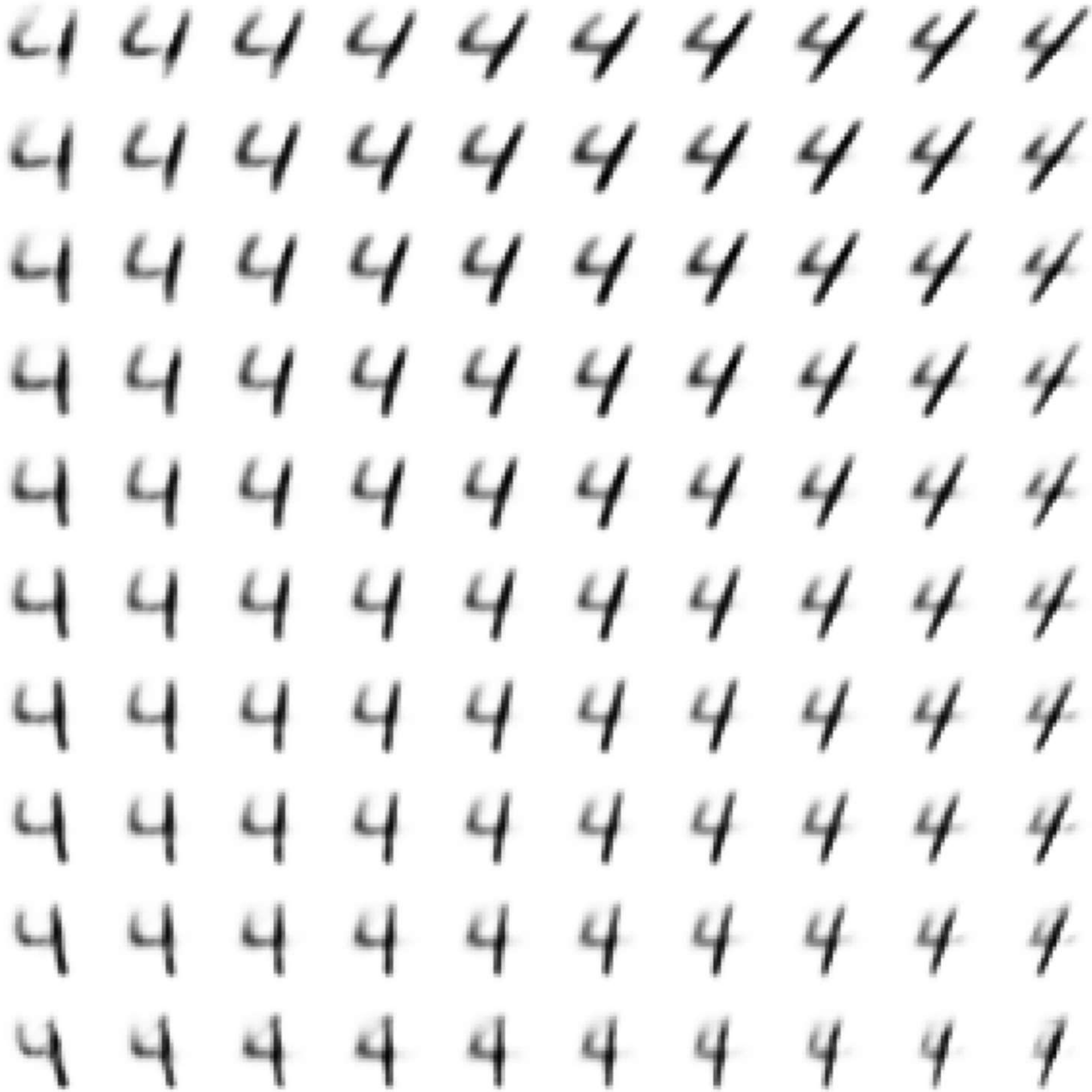
$$q_t(z_i) = p_\theta(z_i \mid x_i)$$

これを求めるのが難しい問題⇒VAE

$$\theta_t = \arg\max_\theta \sum_{i=1}^{n} L[q_t(z_i), \theta; x_i]$$

変分オートエンコーダー
(Variational auto-encoder)

Diederik P Kingma, Max Welling,
Auto-Encoding Variational Bayes , 2014



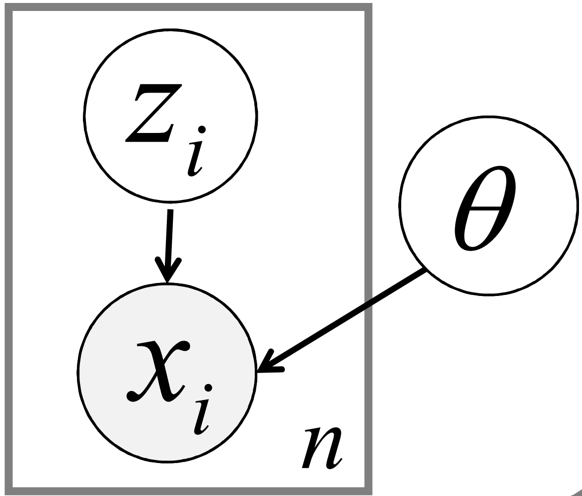
Kingma+, Semi-supervised Learning with Deep Generative Models , NIPS2014

例. ベルヌーイ多層パーセプトロン

$$x_i \in \{0,1\}^D$$

$$z_i \in R^K$$

$$\theta = \{\mathrm{W}_1, \mathrm{W}_2, b_1, b_2\}$$



K次元正規分布

$$z_i \sim N(0, I_K)$$

Element-wiseシグモイド関数

$$\pi_i = \mathrm{sigmoid}(\mathrm{W}_2 \tanh(\mathrm{W}_1 z_i + b_1) + b_2)$$

Element-wise ベルヌーイ分布

$$x_i \sim \mathrm{Be}(\pi_i)$$

Kingma+, Semi-supervised Learning with Deep Generative Models , NIPS2014

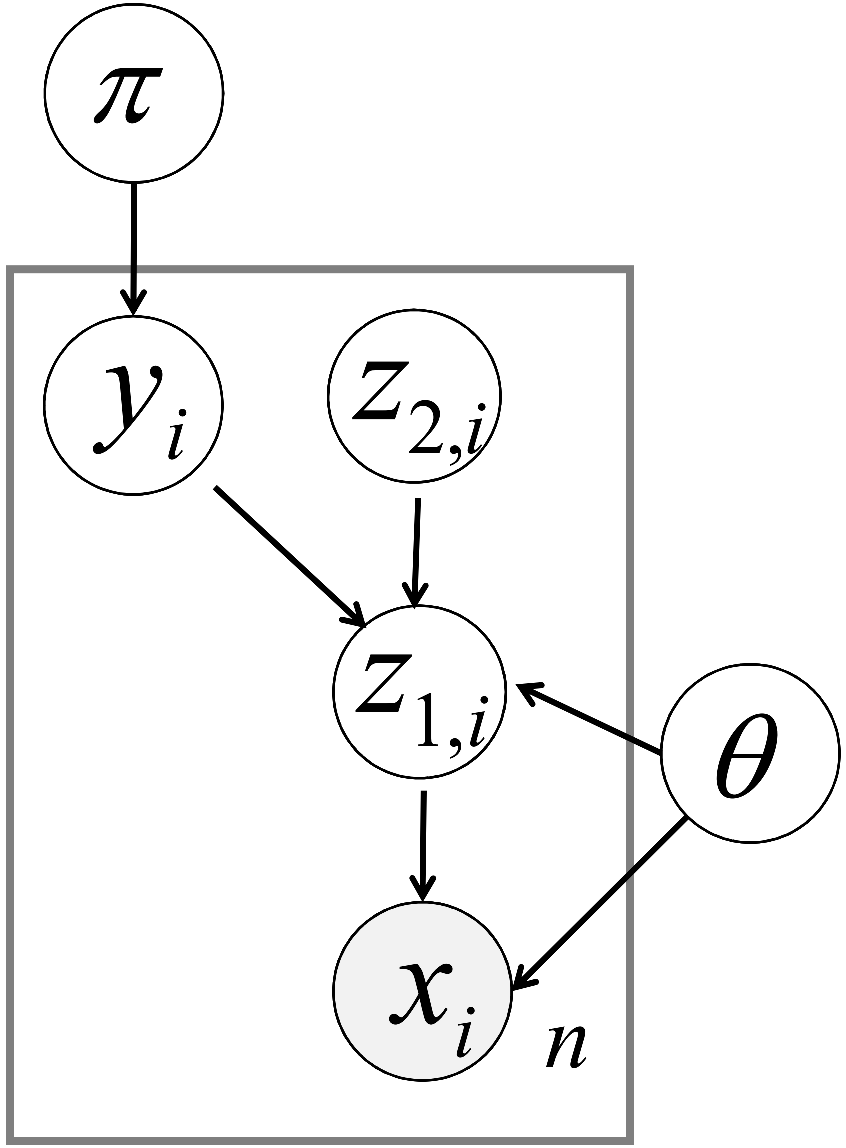
例. 半教師あり生成モデル

$$z_{2,i} \sim N(0, I_K)$$

$$y_i \sim \mathrm{Cat}(\pi)$$

$$z_{1,i} \sim p_\theta(z_1 \mid y_i, z_{2,i})$$

$$x_i \sim p_\theta(x \mid z_{1,i})$$



目的関数

$$\log p_\theta(x_{1:n})$$

$$\geq \sum_{i=1}^{n} \int q(z_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q(z_i)} dz_i$$

$$\equiv \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$ 変分下限などと呼ばれる

交互最適化アルゴリズム

$$q_t(z_i) = p_\theta(z_i \mid x_i)$$

これを求めるのが
難しい問題⇒VAE

$$\theta_t = \arg\max_\theta \sum_{i=1}^{n} L[q_t(z_i), \theta; x_i]$$

目的関数

$$\log p_\theta(x_{1:n})$$

$$\geq \sum_{i=1}^{n} \int q(z_i) \log \frac{p_\theta(x_i \mid z_i)\, p(z_i)}{q(z_i)} dz_i$$

$$\equiv \sum_{i=1}^{n} L[q(z_i), \theta; x_i]$$ 変分下限などと呼ばれる

交互最適化アルゴリズム

$$q_t(z_i) = \underset{q(z_i)}{\arg\max}\, L[q(z_i), \theta_t; x_i]$$

解けるクラスに限定

$$\theta_t = \underset{\theta}{\arg\max} \sum_{i=1}^{n} L[q_t(z_i), \theta; x_i]$$

変分近似事後分布

$$q(z_i) = q_\phi(z_i \mid x_i) = \prod_{k=1}^{K} q_\phi(z_{i,k} \mid x_i)$$

$$= \prod_{k=1}^{K} N(z_{i,k} \mid \mu_k(x_i), \sigma_k^2(x_i))$$

$$\mu_k(x_i) = \tanh(\mathrm{W}_2^\mu \tanh(\mathrm{W}_1^\mu x_i + b_1^\mu) + b_2^\mu)$$

$$\sigma_k^2(x_i) = \exp(\mathrm{W}_2^\sigma \tanh(\mathrm{W}_1^\sigma x_i + b_1^\sigma) + b_2^\sigma)$$

$$\phi = \{W_1^\mu, W_2^\sigma, W_1^\mu, W_2^\mu, b_1^\mu, b_2^\mu, b_1^\sigma, b_2^\sigma\}$$

目的関数

$$\log p_\theta(x_i)$$

$$\geq \int q_\phi(z_i \mid x_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q_\phi(z_i \mid x_i)} dz_i$$

$$\equiv L[\varphi, \theta; x_i]$$

交互最適化アルゴリズム

$$\phi_t = \arg\max_{\phi} \sum_{i=1}^{n} L[\phi, \theta_t; x_i]$$

$$\theta_t = \arg\max_{\theta} \sum_{i=1}^{n} L[\phi_t, \theta; x_i]$$

目的関数

$$\underline{\log p_\theta(x_i)}$$

$$\geq \int q_\phi(z_i \mid x_i) \log \frac{p_\theta(x_i \mid z_i) p(z_i)}{q_\phi(z_i \mid x_i)} dz_i$$

$$\equiv L[\varphi, \theta; x_i]$$

交互最適化アルゴリズム(勾配法)

$$\phi_t = \phi_{t-1} + \rho_t \partial_\phi \sum_{i=1}^{n} L[\phi, \theta_t; x_i]$$

$$\theta_t = \theta_{t-1} + \rho_t \partial_\theta \sum_{i=1}^{n} L[\phi_t, \theta; x_i]$$

変分下限の分析

$$L[\varphi,\theta;x_i] =$$

$$\int q_\varphi(z_i \mid x_i)\log p_\theta(x_i \mid z_i)dz_i$$

$$-\int q_\varphi(z_i \mid x_i)\log\frac{q_\varphi(z_i \mid x_i)}{p(z_i)}dz_i$$

変分下限の分析

$$L[\varphi, \theta; x_i] =$$

解析的に解けない

$$\int q_\varphi(z_i \mid x_i) \log p_\theta(x_i \mid z_i) dz_i$$

$$- \int q_\varphi(z_i \mid x_i) \log \frac{q_\varphi(z_i \mid x_i)}{p(z_i)} dz_i$$

KL情報量(Kullback-liber divergence)
解析的に積分計算可能

変分下限の分析

$$q_\phi(z_i \mid x_i) = \prod_{k=1}^{K} N(z_{i,k} \mid \mu_{i,k}, \sigma_{i,k}^2)$$

$$p(z_i) = \prod_{k=1}^{K} N(z_{i,k} \mid 0,1)$$

$$-\int q_\varphi(z_i \mid x_i) \log \frac{q_\varphi(z_i \mid x_i)}{p(z_i)} dz_i$$

$$= \frac{1}{2} \sum_{k=1}^{K} (1 + \log \sigma_{i,k}^2(x_i) - \mu_{i,k}^2(x_i) - \sigma_{i,k}^2(x_i))$$

変分下限の分析

$$L[\varphi,\theta;x_i] =$$

解析的に解けない

$$\int q_\varphi(z_i \mid x_i)\log p_\theta(x_i \mid z_i)dz_i$$

$$-\int q_\varphi(z_i \mid x_i)\log\frac{q_\varphi(z_i \mid x_i)}{p(z_i)}dz_i$$

KL情報量(Kullback-liber divergence)
解析的に積分計算可能

変分下限の分析

$$L[\varphi,\theta;x_i]=$$

解析的な解はない

$$\int q_\varphi(z_i|x_i)\log p_\theta(x_i|z_i)dz_i$$

$$-\int q_\varphi(z_i|x_i)\log\frac{q_\varphi(z_i|x_i)}{p(z_i)}dz_i$$

KL情報量(Kullback-liber divergence)
解析的に積分計算可能

＜積分計算できない場合の対処法＞

サンプリングによる近似

を使えばいいのでは？

変分下限の近似

解析的に解けない

$$\int q_\varphi(z_i \mid x_i) \log p_\theta(x_i \mid z_i) dz_i$$

$$\approx \frac{1}{S} \sum_{s=1}^{S} \log p_\theta(x_i \mid z_i^{(s)})$$

$$z_i^{(s)} \sim q_\varphi(z_i \mid x_i)$$

変分下限の近似

$$L[\varphi, \theta; x_i] \approx$$

$$\frac{1}{S}\sum_{s=1}^{S} \log p_\theta(x_i \mid z_i^{(s)})$$

$$-\int q_\varphi(z_i \mid x_i) \log \frac{q_\varphi(z_i \mid x_i)}{p(z_i)} dz_i$$

変分下限の近似

$$L[\varphi,\theta;x_i] \approx \frac{1}{S}\sum_{s=1}^{S}\log p_\theta(x_i \mid z_i^{(s)})$$

$$\int q_\varphi(z_i \mid x_i)\log\frac{q_\varphi(z_i \mid x_i)}{p(z_i)}dz_i$$

＜積分計算できない場合の対処法＞

サンプリングによる近似

で問題は解決するのか？

求めたかったものは勾配

$$\partial_{\varphi} L[\varphi, \theta; x_i] =$$

$$\partial_{\varphi} \int q_{\varphi}(z_i \mid x_i) \log p_{\theta}(x_i \mid z_i) dz_i$$

$$\partial_{\varphi} \quad -\int q_{\varphi}(z_i \mid x_i) \log \frac{q_{\varphi}(z_i \mid x_i)}{p(z_i)} dz_i$$

変分下限の近似

$$L[\varphi, \theta; x_i] \approx$$

φに関する勾配が計算できない

$$\frac{1}{S}\sum_{s=1}^{S} \log p_\theta(x_i \mid z_i^{(s)})$$

$$-\int q_\varphi(z_i \mid x_i) \log \frac{q_\varphi(z_i \mid x_i)}{p(z_i)} dz_i$$

Reparameterization trick

正規分布からのサンプリング

$$E_{z\sim N(\mu,\sigma^2)}[f(z)] \approx \frac{1}{S}\sum_{s=1}^{S} f(z^{(s)})$$

$$z_i^{(s)} \sim N(\mu,\sigma^2)$$

$$\left\{\begin{array}{l} z_i^{(s)} = \mu + \sigma\xi^{(s)} \\ \xi^{(s)} \sim N(0,1) \end{array}\right.$$

Reparameterization trick

$$E_{z\sim N(\mu,\sigma^2)}[f(z)] \approx \frac{1}{S}\sum_{s=1}^{S} f(\mu+\sigma\xi^{(s)})$$

$$z_i^{(s)} \sim N(\mu,\sigma^2)$$

$$\begin{cases} z_i^{(s)} = \mu + \sigma\xi^{(s)} \\ \xi^{(s)} \sim N(0,1) \end{cases}$$

変分下限の近似

$$L[\varphi,\theta;x_i] \approx$$

$$\frac{1}{S}\sum_{s=1}^{S}\log p_\theta(x_i \mid z_i^{(s)})$$

$$-\int q_\varphi(z_i \mid x_i)\log\frac{q_\varphi(z_i \mid x_i)}{p(z_i)}dz_i$$

変分下限の近似

$$L[\varphi, \theta; x_i] \approx$$

$$\frac{1}{S}\sum_{s=1}^{S} \log p_\theta(x_i \mid z_i^{(s)})$$

$$z_{i,k}^{(s)} = \mu_k(x_i) + \sigma_k(x_i)\xi_{i,k}{}^{(s)}$$

$$\xi_{i,k}^{(s)} \sim N(0,1)$$

変分下限の近似

$$\frac{1}{S}\sum_{s=1}^{S}\log p_\theta(x_i \mid z_i^{(s)})$$

$$= x_{i,k}\log \pi_{i,k}^{(s)} + (1 - x_{i,k})\log(1 - \pi_{i,k}^{(s)})$$

$$\pi_i^{(s)} = \mathrm{sigmoid}(\mathrm{W}_2 \tanh(\mathrm{W}_1\, z_i^{(s)} + b_1) + b_2)$$

ベルヌーイ多層パーセプトロンの変分下限

$$L[\varphi,\theta;x_i] \approx \frac{1}{S}\sum_{s=1}^{S} \log p_\theta(x_i \mid z_i^{(s)})$$

$$\overbrace{x_{i,k} \log \pi_{i,k}^{(s)} + (1-x_{i,k}) \log(1-\pi_{i,k}^{(s)})}$$

$$-\int q_\varphi(z_i \mid x_i) \log \frac{q_\varphi(z_i \mid x_i)}{p(z_i)} dz_i$$

$$\overbrace{\frac{1}{2}\sum_{k=1}^{K}(1+\log\sigma_k^2(x_i)-\mu_k^2(x_i)-\sigma_k^2(x_i))}$$

交互最適化アルゴリズム(勾配法)

$$\phi_t = \phi_{t-1} + \rho_t \partial_\phi \sum_{i=1}^{n} L[\phi, \theta_t; x_i]$$

$$\theta_t = \theta_{t-1} + \rho_t \partial_\theta \sum_{i=1}^{n} L[\phi_t, \theta; x_i]$$

nが大きい

⇒ミニバッチを用いる（確率的最適化）

勾配の計算が難しい

⇒自動微分を用いる

アルゴリズムの意味を考える

変分下限の分析：θを求める視点

$$\arg\max_{\theta} L[\varphi,\theta;x_i]$$

$$= \arg\max_{\theta} \int q_{\varphi}(z_i \mid x_i) \log p_{\theta}(x_i \mid z_i) dz_i$$

$$- \int q_{\varphi}(z_i \mid x_i) \log \frac{q_{\varphi}(z_i \mid x_i)}{p(z_i)} dz_i$$

変分下限の分析：θ を求める視点

$$\arg\max_{\theta} L[\varphi, \theta; x_i]$$

$$\approx \arg\max_{\theta} \frac{1}{S}\sum_{s=1}^{S} \log p_{\theta}(x_i \mid z_i^{(s)})$$

$$z_i^{(s)} \sim q_{\varphi}(z_i \mid x_i)$$

変分下限の分析：θ を求める視点

$$\arg\max_{\theta} L[\varphi, \theta; x_i]$$

$$\approx \arg\min_{\theta} \frac{1}{S}\sum_{s=1}^{S} -\log p_{\theta}(x_i \mid z_i^{(s)})$$

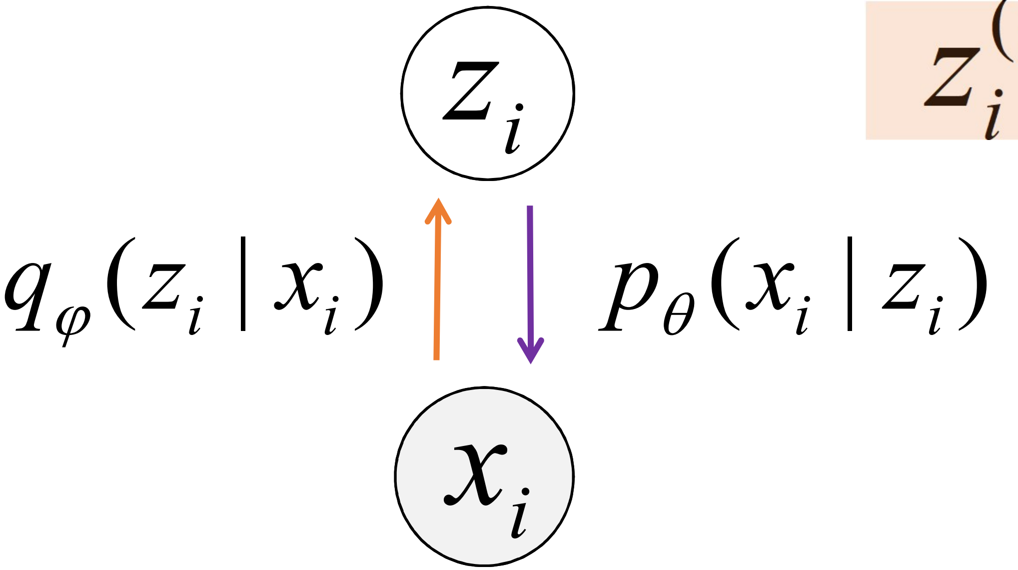
$$z_i^{(s)} \sim q_{\varphi}(z_i \mid x_i)$$

変分下限の分析：θ を求める視点

$$\arg\max_{\theta} L[\varphi, \theta; x_i]$$

デコードエラー

$$\approx \arg\min_{\theta} \frac{1}{S} \sum_{s=1}^{S} -\log p_\theta(x_i \mid z_i^{(s)})$$



$$z_i^{(s)} \sim q_\varphi(z_i \mid x_i)$$

エンコード

変分オートエンコーダ
と呼ばれる